

Soutenance de Stage

Factorisation en Matrices Non-Négatives pour la Reconnaissance Multi-Source Incrémentale

Arnaud Dessein

Encadrants : Arshia Cont, Guillaume Lemaitre

29 juin 2009

Plan

- 1 Etat de l'art
 - Factorisation en matrices non-négatives
 - Contexte de la reconnaissance des sons
 - Reconnaissance multi-source incrémentale
- 2 Contrôle de la parcimonie
- 3 Résultats
- 4 Conclusion

Factorisation en matrices non-négatives

Problème standard de *factorisation en matrices non-négatives* (NMF)

[Lee & Seung, 1999].

- Soit $\mathbf{V} \in \mathbb{R}_+^{n \times m}$ et $r < \min(n, m)$, trouver $\mathbf{W} \in \mathbb{R}_+^{n \times r}$ et $\mathbf{H} \in \mathbb{R}_+^{r \times m}$ telles que :

$$\mathbf{V} \approx \mathbf{WH}$$

- Minimisation par rapport à une *fonction coût* : la distance euclidienne.

Factorisation en matrices non-négatives

Problème standard de *factorisation en matrices non-négatives* (NMF)

[Lee & Seung, 1999].

- Soit $\mathbf{V} \in \mathbb{R}_+^{n \times m}$ et $r < \min(n, m)$, trouver $\mathbf{W} \in \mathbb{R}_+^{n \times r}$ et $\mathbf{H} \in \mathbb{R}_+^{r \times m}$ telles que :

$$\mathbf{V} \approx \mathbf{WH}$$

- Minimisation par rapport à une *fonction coût* : la distance euclidienne.
- Interprétation :

$$\mathbf{v}_j \approx \mathbf{W}\mathbf{h}_j = \sum_i h_{ij} \mathbf{w}_i$$

- \mathbf{w}_i : *vecteurs de base*.
- h_{ij} : *coefficients de décomposition*.

Contexte de la reconnaissance des sons

Rappel du modèle.

$$\mathbf{V} \approx \mathbf{WH}$$

$$\mathbf{v}_j \approx \mathbf{Wh}_j = \sum_i h_{ij} \mathbf{w}_i$$

- Choix commun des paramètres :
 - \mathbf{V} : représentation temps-fréquence.
 - \mathbf{v}_j : trames successives.
 - \mathbf{w}_i : modèles spectraux.
 - h_{ij} : coefficients d'activation.

Contexte de la reconnaissance des sons

Rappel du modèle.

$$\mathbf{V} \approx \mathbf{WH}$$

$$\mathbf{v}_j \approx \mathbf{W}\mathbf{h}_j = \sum_i h_{ij} \mathbf{w}_i$$

- Choix commun des paramètres :
 - \mathbf{V} : représentation temps-fréquence.
 - \mathbf{v}_j : trames successives.
 - \mathbf{w}_i : modèles spectraux.
 - h_{ij} : coefficients d'activation.
- Exemple d'application : transcription de la musique polyphonique [Smaragdis & Brown, 2003, Abdallah & Plumbley, 2004, Raczynski et al., 2007, Vincent et al., 2008, Bertin et al., 2009].

Contexte de la reconnaissance des sons

Rappel du modèle.

$$\mathbf{V} \approx \mathbf{WH}$$

$$\mathbf{v}_j \approx \mathbf{W}\mathbf{h}_j = \sum_i h_{ij} \mathbf{w}_i$$

- Choix commun des paramètres :
 - \mathbf{V} : représentation temps-fréquence.
 - \mathbf{v}_j : trames successives.
 - \mathbf{w}_i : modèles spectraux.
 - h_{ij} : coefficients d'activation.
- Exemple d'application : transcription de la musique polyphonique [Smaragdis & Brown, 2003, Abdallah & Plumbley, 2004, Raczynski et al., 2007, Vincent et al., 2008, Bertin et al., 2009].
- Limites pour notre problème.

Reconnaissance multi-source incrémentale

- Vers une *décomposition non-négative* :
 - 1 Apprendre des modèles w_i avant la décomposition.
 - 2 Regrouper ces modèles dans \mathbf{W} fixe pendant la décomposition.
 - 3 Résoudre un problème NMF pour chaque trame :

$$\mathbf{v} \approx \mathbf{W}\mathbf{h}$$

Reconnaissance multi-source incrémentale

- Vers une *décomposition non-négative* :
 - 1 Apprendre des modèles \mathbf{w}_i avant la décomposition.
 - 2 Regrouper ces modèles dans \mathbf{W} fixe pendant la décomposition.
 - 3 Résoudre un problème NMF pour chaque trame :

$$\mathbf{v} \approx \mathbf{W}\mathbf{h}$$

- Applications :
 - Analyse de la parole [Sha & Saul, 2005].
 - Transcription [Paulus & Virtanen, 2005, Niedermayer, 2008].
 - Suivi de partition [Cont, 2006].
 - Reconnaissance multi-f0 et multi-instrument [Cont et al., 2007].
 - Evaluation de la lecture à vue [Cheng et al., 2008].

Reconnaissance multi-source incrémentale

- Vers une *décomposition non-négative* :
 - 1 Apprendre des modèles \mathbf{w}_i avant la décomposition.
 - 2 Regrouper ces modèles dans \mathbf{W} fixe pendant la décomposition.
 - 3 Résoudre un problème NMF pour chaque trame :

$$\mathbf{v} \approx \mathbf{W}\mathbf{h}$$

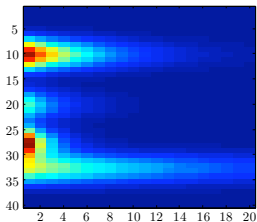
- Applications :
 - Analyse de la parole [Sha & Saul, 2005].
 - Transcription [Paulus & Virtanen, 2005, Niedermayer, 2008].
 - Suivi de partition [Cont, 2006].
 - Reconnaissance multi-f0 et multi-instrument [Cont et al., 2007].
 - Evaluation de la lecture à vue [Cheng et al., 2008].
- Problème majeur : contrôler la *parcimonie* de la décomposition [Cont, 2006, Cont et al., 2007].

Plan

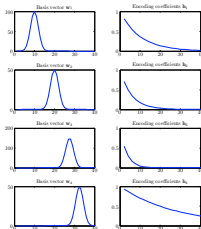
- 1 Etat de l'art
- 2 **Contrôle de la parcimonie**
 - Illustration et motivations
 - Optimisation par gradient projeté
 - Programmations conique de second ordre et quadratique convexe
- 3 Résultats
- 4 Conclusion

Illustration et motivations

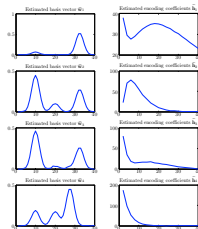
- Expérience de [Paatero, 1997].



(a) $V = WH + |N|$.



(b) W et H .



(c) \hat{W} et \hat{H} .

FIGURE: Expérience de Paatero avec NMF.

Illustration et motivations

- Expérience de [Paatero, 1997].

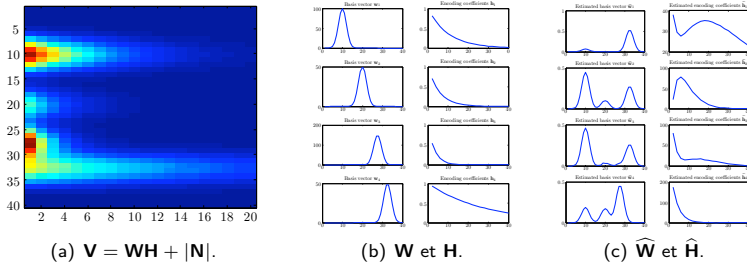


FIGURE: Expérience de Paatero avec NMF.

- Vers l'utilisation de techniques d'*optimisation convexe* [Boyd & Vandenberghe, 2004].

Optimisation par gradient projeté

- Utilisation de l'*optimisation par gradient projeté* par [Hoyer, 2004].
- Mesure de parcimonie :

$$\text{sp}(\mathbf{x}) = \frac{\sqrt{n} - \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2}{\sqrt{n} - 1}$$

$$0 \leq \text{sp}(\mathbf{x}) \leq 1$$

Optimisation par gradient projeté

- Utilisation de l'*optimisation par gradient projeté* par [Hoyer, 2004].
- Mesure de parcimonie :

$$\text{sp}(\mathbf{x}) = \frac{\sqrt{n} - \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2}{\sqrt{n} - 1}$$

$$0 \leq \text{sp}(\mathbf{x}) \leq 1$$

- Choix d'une parcimonie s par l'utilisateur.
- Descente de gradient avec étape de projection sur le cône de parcimonie s .

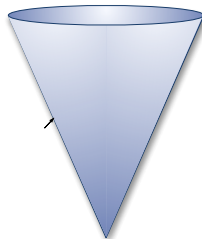


FIGURE: Projection sur un cône de parcimonie.

Programmations conique de second ordre et quadratique convexe

- Utilisation de la *programmation conique de second ordre* par [Heiler & Schnörr, 2006].
- Optimisation entre deux cônes de parcimonie $s_{min} \leq \text{sp}(\mathbf{x}) \leq s_{max}$.

Programmations conique de second ordre et quadratique convexe

- Utilisation de la *programmation conique de second ordre* par [Heiler & Schnörr, 2006].
- Optimisation entre deux cônes de parcimonie $s_{min} \leq \text{sp}(\mathbf{x}) \leq s_{max}$.
- Problème convexe retourné assez complexe.
- Simplification par une séquence de programmes coniques de second ordre.

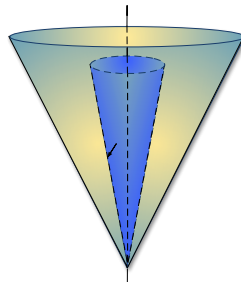


FIGURE: Optimisation entre deux cônes de parcimonie.

Programmations conique de second ordre et quadratique convexe

- Utilisation de la *programmation conique de second ordre* par [Heiler & Schnörr, 2006].
- Optimisation entre deux cônes de parcimonie $s_{min} \leq \text{sp}(\mathbf{x}) \leq s_{max}$.
- Problème convexe retourné assez complexe.
- Simplification par une séquence de programmes coniques de second ordre.
- Utilisation de la *programmation quadratique convexe* [Boyd & Vandenberghe, 2004].
- Plus efficace et robuste.
- Introduction de *pénalités* dans la fonction coût.

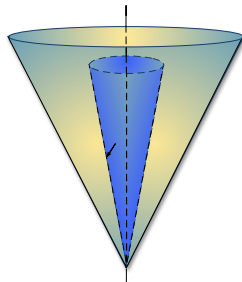
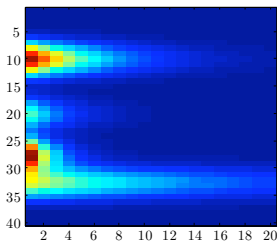


FIGURE: Optimisation entre deux cônes de parcimonie.

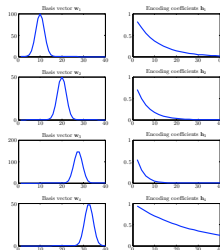
Plan

- 1 Etat de l'art
- 2 Contrôle de la parcimonie
- 3 Résultats**
 - Expérience de Paatero
 - Reconnaissance multi-f0 et multi-instrument
 - Analyse de scènes sonores complexes
- 4 Conclusion

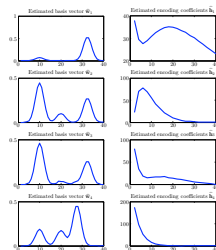
Expérience de Paatero



(a) $\mathbf{V} = \mathbf{W}\mathbf{H} + |\mathbf{N}|$.



(b) \mathbf{W} et \mathbf{H} .



(c) $\widehat{\mathbf{W}}$ et $\widehat{\mathbf{H}}$.

FIGURE: Expérience de Paatero avec NMF et SNMF.

Expérience de Paatero

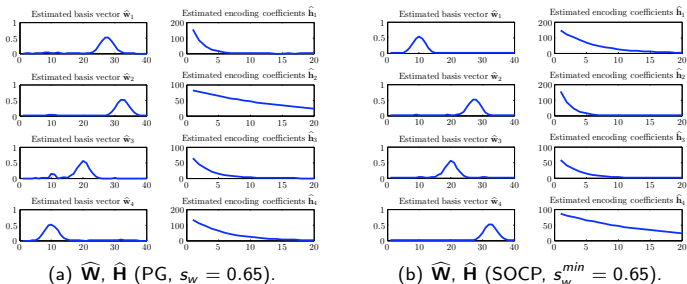


FIGURE: Expérience de Paatero avec NMF et SNMF.

Reconnaissance multi-f0 et multi-instrument

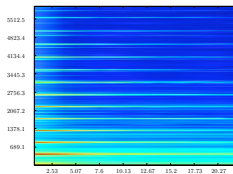
- Représentation :
 - Le *spectre de modulation* [Sukittanon et al., 2004].
 - Non-négativité, additivité [Atlas & Janssen, 2005].
 - Déjà utilisé, discriminatif [Cont et al., 2007].

Reconnaissance multi-f0 et multi-instrument

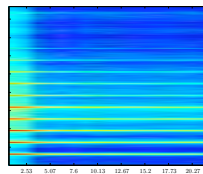
- Représentation :
 - Le *spectre de modulation* [Sukittanon et al., 2004].
 - Non-négativité, additivité [Atlas & Janssen, 2005].
 - Déjà utilisé, discriminatif [Cont et al., 2007].
- Apprentissage des modèles pour chaque note de chaque instrument avec NMF.

Reconnaissance multi-f0 et multi-instrument

- Représentation :
 - Le *spectre de modulation* [Sukittanon et al., 2004].
 - Non-négativité, additivité [Atlas & Janssen, 2005].
 - Déjà utilisé, discriminatif [Cont et al., 2007].
- Apprentissage des modèles pour chaque note de chaque instrument avec NMF.



(a) A4 du piano.



(b) A4 de la flute.

FIGURE: Modèles de A4 pour la flute et le piano.

Reconnaissance multi-f0 et multi-instrument

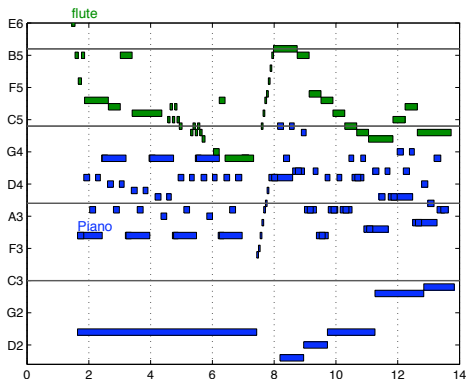
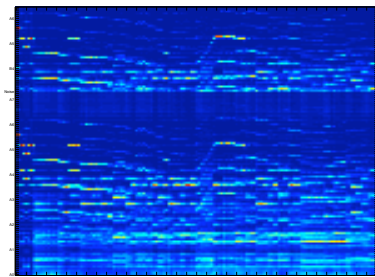


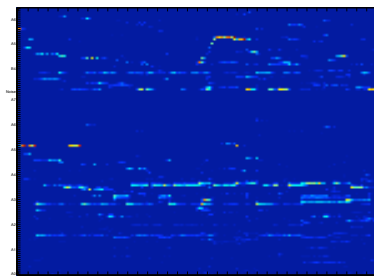
FIGURE: Piano-roll de la *Sonate pour Flute et Piano* de Poulenc.

Reconnaissance multi-f0 et multi-instrument

- Evaluation subjective.



(a) Algorithme de [Cont et al., 2007].



(b) Algorithme CQP.

FIGURE: Coefficients d'activation \mathbf{H} obtenus.

- Evaluation objective en cours.

Analyse de scènes sonores complexes

- Spectre d'amplitude.
- Apprentissage des modèles.

Analyse de scènes sonores complexes

- Spectre d'amplitude.
- Apprentissage des modèles.
- Création d'une scène sonore.

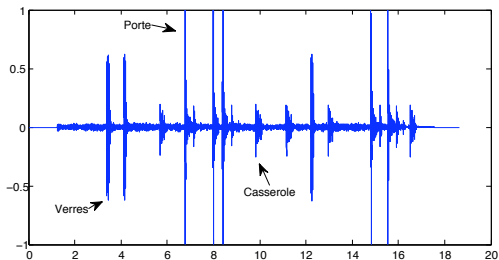
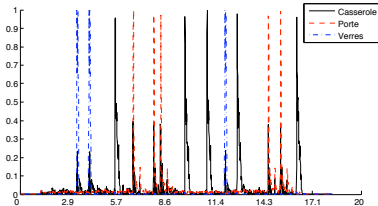


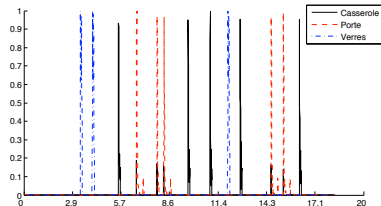
FIGURE: Forme d'onde de la scène sonore.

Analyse de scènes sonores complexes

- Séquence : V – V – C – P – P – P – C – C – V – C – P – P – C.



(a) Coefficients d'activation avec ND.



(b) Coefficients d'activation avec SND.

FIGURE: Analyse d'une scène sonore complexe.

Plan

- 1 Etat de l'art
- 2 Contrôle de la parcimonie
- 3 Résultats
- 4 Conclusion**

Conclusion

Résumé du travail et perspectives.

- Espace et géométrie du problème.
- Représentations.
- Temporalité des évènements.
- Pouvoir de généralisation et robustesse.

Conclusion

Résumé du travail et perspectives.

- Espace et géométrie du problème.
 - Représentations.
 - Temporalité des évènements.
 - Pouvoir de généralisation et robustesse.
-
- La distance euclidienne avec interprétation géométrique.
 - Autres fonctions coûts, divergences.
 - La *divergence d'Itakura-Saito* avec interprétation bayésienne [Févotte et al., 2009].

Conclusion

Résumé du travail et perspectives.

- Espace et géométrie du problème.
 - Représentations.
 - Temporalité des évènements.
 - Pouvoir de généralisation et robustesse.
-
- Transformée en ondelettes.
 - Représentations complexes pour \mathbf{V} et \mathbf{W} .
 - Tenseurs [Welling & Weber, 2001, Cichocki & Zdunek, 2006, Friedlander, 2006].

Conclusion

Résumé du travail et perspectives.

- Espace et géométrie du problème.
 - Représentations.
 - Temporalité des évènements.
 - Pouvoir de généralisation et robustesse.
-
- Utilisation du spectre de modulation.
 - Extension du modèle NMF [Smaragdis, 2004].
 - Représentation par états.

Conclusion

Résumé du travail et perspectives.

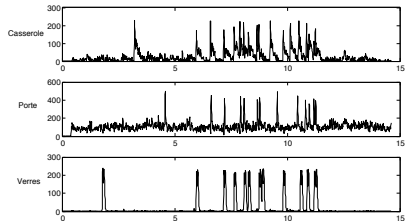
- Espace et géométrie du problème.
 - Représentations.
 - Temporalité des évènements.
 - Pouvoir de généralisation et robustesse.
-
- **W** fixe.
 - Relâchement de la contrainte sur **W**.
 - Contrôle explicite de la robustesse.

Questions

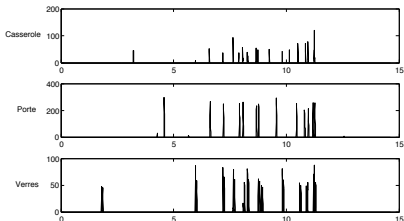
Merci pour votre attention !



Analyse d'une scène complexe



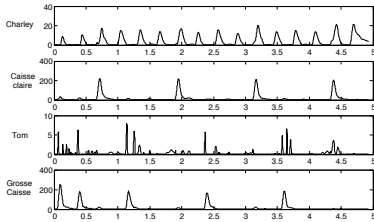
(a) Coefficients d'activation avec ND.



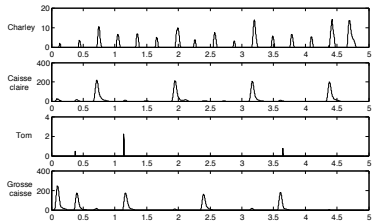
(b) Coefficients d'activation avec SND.

FIGURE: Analyse d'une scène complexe.

Rythme de batterie



(a) Coefficients d'activation avec ND.



(b) Coefficients d'activation avec SND.

FIGURE: Rythme de batterie.

Bibliographie I



Abdallah, S. A. & Plumbley, M. D. (2004).

Polyphonic music transcription by non-negative sparse coding of power spectra.
In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)* (pp. 318–325). Barcelona, Spain.



Atlas, L. E. & Janssen, C. (2005).

Coherent modulation spectral filtering for single-channel music source separation.
In *IEEE International Conference in Acoustics and Speech Signal Processing (ICASSP)*.



Bertin, N., Badeau, R., & Vincent, E. (2009).

Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription.
Technical report, TELECOM ParisTech.



Boyd, S. & Vandenberghe, L. (2004).

Convex Optimization.
Cambridge University Press.



Cheng, C.-C., Hu, D. J., & Saul, L. K. (2008).

Nonnegative matrix factorization for real time musical analysis and sight-reading evaluation.
In *Acoustics, Speech and Signal Processing 2008, IEEE International Conference on* (pp. 2017–2020). Las Vegas, NV, USA.



Cichocki, A. & Zdunek, R. (2006).

NMFLAB/NTFLAB – MATLAB Toolbox for Non-Negative Matrix/Tensor Factorization.
<http://www.bsp.brain.riken.jp/ICALAB/nmflab.html>.



Cont, A. (2006).

Realtime multiple pitch observation using sparse non-negative constraints.
In *International Symposium on Music Information Retrieval (ISMIR)* Victoria, Canada.

Bibliographie II



Cont, A., Dubnov, S., & Wessel, D. (2007).

Realtime multiple-pitch and multiple-instrument recognition for music signals using sparse non-negative constraints.
In Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07) Bordeaux, France.



Friedlander, M. P. (2006).

BCLS : A large-scale solver for bound-constrained least squares.
<http://www.cs.ubc.ca/~mpf/bcls/>.



Févotte, C., Bertin, N., & Durrieu, J.-L. (2009).

Nonnegative matrix factorization with the itakura-saito divergence : With application to music analysis.
Neural Computation, 21(3), 793–830.



Heiler, M. & Schnörr, C. (2006).

Learning sparse representations by non-negative matrix factorization and sequential cone programming.
Journal of Machine Learning Research, 7, 1385–1407.



Hoyer, P. O. (2004).

Non-negative matrix factorization with sparseness constraints.
Journal of Machine Learning Research, 5, 1457–1469.



Lee, D. D. & Seung, H. S. (1999).

Learning the parts of objects by non-negative matrix factorization.
Nature, 401(6755), 788–791.



Niedermayer (2008).

Non-negative matrix division for the automatic transcription of polyphonic music.
In Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR) (pp. 544–549).

Bibliographie III



Paatero, P. (1997).

Least squares formulation of robust non-negative factor analysis.
Chenometrics and Intelligent Laboratory Systems, 37(1), 23–35.



Paulus, J. & Virtanen, T. (2005).

Drum transcription with non-negative spectrogram factorisation.
In *Proceedings of the 13th European Signal Processing Conference (EUSIPCO 05)* Antalya, Turkey.



Raczyński, S. A., Ono, N., & Sagayama, S. (2007).

Harmonic nonnegative matrix approximation for multipitch analysis of musical sounds.
In *Proceedings of ASJ Autumn Meeting* (pp. 827–830).



Sha, F. & Saul, L. K. (2005).

Real-time pitch determination of one or more voices by nonnegative matrix factorization.
Advances in Neural Information Processing Systems, 17, 1233–1240.



Smaragdis, P. (2004).

Non-negative matrix factor deconvolution ; extraction of multiple sound sources from monophonic inputs.
In Springer (Ed.), *Proceedings of the 5th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 04)*, volume 3195 of *Lecture Notes in Computer Science* (pp. 494–499). Granada, Spain.



Smaragdis, P. & Brown, J. C. (2003).

Non-negative matrix factorization for polyphonic music transcription.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 177–180). New Paltz, NY.



Sukittanon, S., Atlas, L. E., & Pitton, J. W. (2004).

Modulation-scale analysis for content identification.
IEEE Transactions on Signal Processing, 52(10), 3023–3035.

Bibliographie IV



Vincent, E., Bertin, N., & Badeau, R. (2008).

Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription.
In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on (pp. 109 –112).



Welling, M. & Weber, M. (2001).

Positive tensor factorization.
Pattern Recognition Letters, 22(12), 1255–1261.